A model for spatial multicriteria hierarchical clustering

Karim Lidouh* and Yves De Smet

Department of Computer and Decision Engineering, SMG Research Unit, École polytechnique de Bruxelles, Université libre de Bruxelles, 50 Avenue F.D. Roosevelt CP 210/01, 1050 Brussels, Belgium Email: klioud@ulb.ac.be Email: yvdesmet@ulb.ac.be *Corresponding author

Abstract: Research on the problem of multicriteria territory partitioning is at its begin. This is mainly due to the fact that it involves tools from fields that are to this day still young. To answer this shortage, we propose an adaptation of a multicriteria clustering method that takes spatial constraints into account. Two variants are described and tested on an illustrative case. This example deals with the partitioning of the Walloon region in Belgium into clusters with a similar level of well-being as perceived by its inhabitants.

Keywords: multiple criteria analysis; hierarchical clustering; preference modelling; territory partitioning.


Biographical notes: Karim Lidouh holds a degree as a Civil Engineer in Computer Sciences, Master in Management and completed his PhD thesis on the integration of multicriteria tools in geographical information systems in 2014. He works as a Teaching Assistant in the fields of statistics and quantitative methods at the Solvay Brussels School of Economics and Management (SBS-EM) of the Université libre de Bruxelles. He also gives part-time courses on operations research and optimisation methods at the IESEG School of Management.

Yves De Smet is an Associate Professor at the Engineering Faculty of the Université libre de Bruxelles. He is the Head of the CoDE-SMG Research Unit. He holds a degree in Mathematics (1998) and PhD in Applied Sciences (2005). His research interests are focused on multicriteria decision aid and multi-objective optimisation. Besides his academic activities he has been involved in different industrial projects. Since 2010, he has been a co-founder of the Decision Sights Spin-off.

This paper is a revised and expanded version of a paper entitled [title] presented at [name, location and date of conference].

Copyright 20XX Inderscience Enterprises Ltd.
1 Introduction

Some of the first cases to have been studied using multicriteria methods about 30 years ago involved spatial entities (Bertier and de Montgolfier, 1978). These have shown the advantages of working with the multicriteria paradigm when dealing with problems that involve spatial components (Roy, 1996). However as can be seen in recent works (Lidouh, 2013), the integration between multicriteria decision aid and geographical information science has experienced a very slow evolution.

This paper will focus on a particular type of spatial problem which is territory partitioning or districting. From a methodological point of view, this problem can be seen as a clustering problem with additional constraints (Zopounidis and Doumpos, 2002), namely that only connected areas can be grouped in the same class or cluster. And even though the literature abounds with cases of territory partitioning (see Ricca et al., 2013), very few of them actually take explicitly multicriteria information into account. One of the sole examples is demonstrated by the work of Tavares-Pereira et al. (2007), where the multicriteria profiles of all the areas to be clustered were taken into account and partitions where generated using a genetic algorithm. To our knowledge, aside from that contribution, very few others proposed significant works that tackle this particular problem.

To the best of our knowledge, De Smet and Guzmán (2004) were the first authors who have considered the extension of clustering algorithms into a multicriteria context. The aim of their contribution was to detect nominal clusters based on a k-means-like approach. This was later extended by De Smet and Eppe (2009) in the context of partially ordered clustering. Since then, several works have been dedicated to this emerging research field (Cailloux et al., 2007; De Smet et al., 2012; De Smet, 2013; Eppe et al., 2014; Meyer and Olteanu, 2013; Rocha et al., 2013; Rocha and Dias, 2013).

This contribution is at the intersection between three domains: clustering, spatial analysis and multicriteria decision aid. The method developed in De Smet and Guzmán (2004) cannot be directly applied to spatial analysis due to the geographic nature of the problems. Therefore it is extended in Section 2 where two main novelties are introduced:

1 Due to spatial constraints, only neighbour entities can be grouped in a given cluster. As a consequence the implementation of the algorithm is more complex since we have to keep track of entities, or groups of entities, that can be merged or not (due to connectivity constraints). Moreover, let us stress that these conditions are likely to evolve during the algorithm execution.

2 We decided to use an ascending hierarchical clustering algorithm (instead of k-means like procedure). A similar approach was used by Rocha and Dias (2013). This choice seems to be more adapted to the spatial constraint mentioned in the previous point. Indeed, when two entities belong to the same group at iteration $k$, they will remain so at iteration $(k + 1)$. On the one hand, this consistency membership seems to be more appropriate to spatial problems. On the other hand, it is easier to control the set of neighbour entities during the algorithm execution. Finally, using a hierarchical clustering algorithm allows to build a complete dendogram that can easily be explored for different partition sizes.
In Section 2 we present the model we use and explain its differences with the initial clustering method. In Section 3, we put this model to use on an illustrative case which studies the well-being in the Walloon region of Belgium. Two maps are proposed with the two variants we have developed. Finally, Section 4, concludes this paper and presents some perspectives.

2 Model

As previously stated, the model we use is based on an earlier article by De Smet and Guzmán (2004). In that first version, the authors proposed an extension of the well-known $k$-means algorithm to multicriteria clustering problems. The result was a nominal clustering that was based on four preference relations. Those allow to characterise the profile of each action, i.e., its relative ‘preferential position’ with respect to the whole dataset. The model was based on the idea that all the actions in a cluster should have similar behaviours in terms of preference, indifference, and incomparability relations. The algorithm thus adopts an approach similar to the $k$-means method (MacQueen et al., 1967; Anderberg, 1973) and stops when the cluster memberships no longer change.

Our approach differs from the previous one in that it is based on a hierarchical procedure. At each step of the process all actions are compared and the pair of actions with the smallest distance between profiles gets merged into a single cluster with a resulting profile. Furthermore, since we are applying this to territory partitioning problems, comparisons are only made between neighbouring actions. This ensures that only connected actions or clusters are merged. After a number of steps equal to the number of actions, all actions are merged together into a single cluster.

In this section, the set of actions to be clustered will be denoted $A = \{a_1, a_2, \ldots, a_n\}$ (where $n$ denotes the number of actions). The set of actions merged together at each given step of the method will be referred to as $\mathcal{B} = \{B_1, B_2, \ldots, B_\mathcal{N}\}$ where $\mathcal{N}$ indicates the current number of clusters ($n$ at the start of the algorithm, and 1 at the end). The information on neighbouring actions will be stored in an $n \times n$ adjacency matrix denoted $C = (c_{ij})$ with $i, j = 1, 2, \ldots, n$ and $c_{ij} = 1$ if $a_i$ and $a_j$ are spatially adjacent. $c_{ij} = 0$ otherwise.

2.1 Profiles

As in De Smet and Guzmán (2004), the profiles of each action will be built using the traditional $(P, I, J)$ (Preference, Indifference, and Incomparability) relations (Vince, 1992). Each action $a_i$’s profile will be a 4-uple $(J(a_i), P^-(a_i), I(a_i), P^+(a_i))$ where:

- $J(a_i) = \{a_j \in A | a_iJa_j\} = P_3(a_i)$
- $P^-(a_i) = \{a_j \in A | a_jPa_i\} = P_2(a_i)$
- $I(a_i) = \{a_j \in A | a_iJa_j\} = P_3(a_i)$
- $P^+(a_i) = \{a_j \in A | a_iPa_j\} = P_4(a_i)$.

We refer the interested reader to the article by De Smet and Guzmán (2004) for a detailed definition of the preference structure we use.
K. Lidouh and Y. De Smet

For practical reasons, these profiles will be stored in a \( n \times 4n \) binary matrix defined as follows:

\[
P = \begin{pmatrix}
P_1(a_1), P_2(a_1), P_3(a_1), P_4(a_1) \\
\ldots \\
P_1(a_n), P_2(a_n), P_3(a_n), P_4(a_n) \\
\end{pmatrix} = \begin{pmatrix}
J(a_1), P^-(a_1), I(a_1), P^+(a_1) \\
\ldots \\
J(a_n), P^-(a_n), I(a_n), P^+(a_n)
\end{pmatrix} = (p_{ij})_{n \times 4n}
\]

where each element is equal to 1 if the corresponding relation between the two actions is true, and equal to 0 otherwise. Elements of this matrix will be noted \( p_{ij} \).

### 2.2 Distance

In order to determine which actions or clusters are to be merged together at each step, it is necessary to compute a distance between their profiles. As these are only composed of 0 and 1 values in the \( P \) matrix, this is easily done using the following equation:

\[
d(B_i, B_j) = 1 - \frac{1}{n^4} \sum_{l=1}^{4} |P_l(B_i) \cap P_l(B_j)| = 1 - \frac{1}{n^4} \sum_{m=1}^{4n} (p_{im} \cdot p_{jm})
\]

In doing so, two clusters will be considered close the more their profiles are alike. The distance between them would then be closer to 0 than 1.

### 2.3 Construction of resulting profiles

There are several ways to construct a resulting profile for a cluster made of a set of actions. We propose two versions for this algorithm.

#### 2.3.1 Voting procedure

The initial version of the multicriteria clustering algorithm (De Smet and Guzmán, 2004) featured only one way of constructing a resulting profile when actions are merged into a single cluster. If \( \{a_{i_1}, a_{i_2}, \ldots, a_{i_p}\} \) are the \( p \) actions in cluster \( B_i \), its profile \( P(B_i) \) was determined using a voting procedure:

\[
a_j \in P_k(B_i) \iff k = \arg \max \sum_{a_{i_1}} \{a_j \in P_k(a_{i_1})
\]

where \( a_{i_1} \) are the actions belonging to the considered cluster, \( k \) is a profile, and \( \{a_j \in P_k(a_{i_1})\} \) is the indicator function (equal to 1 when the action is part of the set, and 0 otherwise). When several values of \( k \) satisfy the condition, the final value is chosen randomly.

This voting procedure allows to obtain a resulting profile that matches the profiles within the cluster as closely as possible. When several profiles can be used, one of them is selected randomly. A direct consequence of this is that the results expected from this procedure will not always be consistent as is shown in Section 3.
2.3.2 Intersection

The second procedure we propose consists in replacing the set of profiles in a cluster with their intersection. The idea behind this approach is that we only keep the common part between all the profiles. Using the same notations as before, we have:

\[ a_j \in P_k(B_i) \iff k = \bigcap_{a_i} P_k(a_i) \]

This approach ensures that the results are always the same. However, this generates incomplete profiles which lead us to distances that become greater after less steps. Therefore, if the number of actions is great, at some point, the associations might no longer make any sense. The smallest distance found could therefore be used as a termination criterion. Indeed, one could for example stop the algorithm as soon as the smallest distance becomes equal to 1, meaning that all the cluster profiles left no longer present any similarities.

2.4 Algorithm

Algorithm 1 shows all the steps of the method as we implemented it. The only part that can be adapted is the construction of the resulting profiles at each iteration that we described at the end of the previous section.

When using a voting procedure, we encounter the same drawbacks as with the k-means method. This is common for such approaches for which there is no uniqueness of results due to the starting conditions and the construction of prototypes for the clusters. Profiles built using a voting procedure will indeed be set randomly when there is an equal number of votes for two types of relations. However, since our approach is hierarchical, this effect is limited. It is also further limited due to the fact that we only merge actions, or groups of actions, that are adjacent.

Algorithm 1 Spatial multicriteria hierarchical clustering

```plaintext
for all numbers of clusters \( n^* \) do
    for all clusters \( B_i \) in \( \mathcal{B} \) do
        Compute the smallest distance to its adjacent neighbours
    end for
    Select the cluster \( B_i \) with the smallest distance \( d(B_i, B_j) \) and assign its neighbour \( B_j \) to it
    Update the set \( \mathcal{B} \) by removing cluster \( B_j \)
    Update matrix \( C \) by removing row \( j \) and replacing row \( i \) with the max of both rows (apply the same to the columns \( i \) and \( j \))
    Update matrix \( P \) by removing row \( j \) and replacing row \( i \) by the resulting profile for the new cluster (apply the same to the columns for each relation \( P_k \))
end for
```

3 Illustrative case: the ICWB in Wallonia

In order to illustrate this method, we chose to use a recently published study from the Walloon Institute for Evaluation, Prospective, and Statistics (IWEPS): the index
of conditions of well-being in Wallonia (ICWB) (Charlier et al., 2014). This index was constructed out of 58 indicators that were grouped in 19 dimensions and then eight families. All of them were evaluated on the 262 municipalities that constitute the Walloon region in Belgium. Figure 1 shows the hierarchical structure of the ICWB. The data for the 19 dimensions was extracted from the report that was published by the IWEPS institute. As the table is quite big \(262 \times 19\) we decided not to include it in this paper, but the original report is freely available on the IWEPS website.

**Figure 1** Hierarchical structure for the ICWB

- **Essential resources**
  - Health, Housing, Education, Employment, Income, Mobility
- **Environment**
  - Environment, Proximity, Safety
- **Relations institutions**
  - Communication, Public management, Democracy
- **Personal relations**
  - Family
- **Social balances**
  - Access to health, Access to employment, Access to income
- **Personal balances**
  - Time management
- **Feelings of well-being**
  - Being happy
- **Values/attitudes**
  - Commitment

*Source: Based on Charlier et al. (2014)*

### 3.1 Preference structure

Instead of applying a standardisation similar to the one used by the IWEPS researchers, we chose to use preference functions inspired from the PROMETHEE methodology (Brans and Vincke, 1985). For each criterion, we defined a preference function where both preference and indifference thresholds were both equal to half of the greatest difference between evaluations (see Figure 2). This ensured that we would obtain diversified profiles with some common parts. Using these functions for each pair of actions, we counted the number of criteria on which an action \(a_i\) is preferred to another \(a_j\) and the number of criteria where the opposite relation was observed:

- if both numbers were different from zero, the actions were considered incomparable \((a_j I a_j)\)
- if both numbers were equal to zero, the actions were considered indifferent \((a_i I a_j)\)
- if only one of the numbers was equal to zero, the better action was preferred to the other \((a_i P a_j\) or \(a_j P a_i))\).

By doing this, we considered equal weights for all the criteria. Let us point out that a description of detailed and justified preferences goes beyond the scope of this paper since it is only used for illustration purposes. This is why we rely on a simplified model.
We also considered two separate sets of criteria: on the one hand the 19 dimensions and on the other hand the eight aggregated families. This allowed us to see how the algorithm behaved when applied on datasets that significantly differ in size. Finally, in order to compare the results, we stopped the algorithm after 200 steps, leaving us with 62 clusters of variable sizes. In the next subsections, we describe two of the results we obtained using the different methods to construct resulting profiles for the clusters.

3.2 Voting procedure

When applying the voting procedure on such a large problem and for a great number of steps, we immediately see that the voting procedure will tend to easily merge actions and obtain clusters with heterogeneous contents. This can be seen in Figure 3 where we observe that a single cluster (i.e., cluster 5) covers almost the entire map. Since the procedure replaces the actions of a cluster by a single profile, small differences do not matter so much.

The numbers indicate the cluster to which each municipality belongs. The colors represent the ICWB score of each cluster computed by using the average of all the municipalities’ scores. A high score is displayed in red while low scores are displayed in yellow.

As mentioned earlier, each attempt at applying this approach leads to different results due to the randomisation aspect of the procedure. Furthermore the great number of steps increases the likelihood that small differences at the beginning might have a great impact on the rest of the computations.

However, an interesting aspect of this approach is that it identifies outliers quite well. Indeed, those municipalities will usually stay isolated in their own cluster until the last steps of the algorithm. The smallest clusters indicate characteristics of Wallonia that match the study done by IWEPS. For instance, the set of yellow clusters that form a belt in the upper part of the region correspond to the industrial and urban areas where the ICWB values are at their lowest. These outliers seem to be the only constant characteristic of the varied results we obtain when applying the method several times.
3.3 Intersection

When applying the intersection procedure the results are fundamentally different. Indeed, as can be seen in Figure 4, each difference counts as the resulting profiles at each step become more and more different from each other, leading us to several clusters of roughly the same size.
Once again the outliers are present, but this time there is no single cluster that occupies the majority of the map. The variety in profiles is indeed preserved and is shown with the different scores in a map that matches the individual ICWB scores more accurately.

### 3.4 Influence of the size of the problem

The number of criteria used does have a significant influence on the results. As expected, we noticed indeed that with 19 criteria we frequently find pairs of alternatives that are incomparable. This of course can lead to several problems as the profiles are almost completely heterogeneous from the start. This means that an approach based on the intersection of profiles will no longer work. Indeed, only a few steps are enough to start having empty profiles that appear in our set of clusters. From that point, the distances evaluated are always maximal and the next steps of the algorithm have no real meaning.

The voting approach however seems to deal with this difficulty a bit better as it will always produce complete profiles. Nonetheless the results obtained with this method should be treated with care as the clusters themselves might contain actions very different from their resulting profile.

Another point to be mentioned is the impact of weights or lack thereof. Indeed, as this method does not make use of any differentiated weights (e.g., we simply counted the criteria for which an alternative is preferred to another, see Section 3.1), it considers all criteria equally to generate the starting profiles of all actions. This does not constitute a problem as such but we need to be aware of this aspect when giving interpretations on the results. Depending on the values, this feature’s influence might have an even greater impact when combined with a large number of criteria.

As far as performances are concerned, both approaches behave very differently with a problem of this size. The intersection approach finishes calculations on a recent computer in merely two seconds for all 262 steps. The voting procedure however increases the duration of these calculations to 45 seconds on the same machine.

### 3.5 Measures of homogeneity and dissimilarity

In order to assess the quality of the results obtained one can make use of the distance between profiles that we defined in Section 2.2. Indeed, by computing the maximum distance between the actions of each cluster, we can get a sense of the homogeneity within the elements of the partition obtained. We did this for the 62 clusters obtained after 200 iterations of both approaches intersection operator and voting procedure). The results of those computations are displayed in Figure 5.

The first thing to be noted is the number of clusters for which the maximum distance is null. For the intersection-based approach, this number is 15 (24% of the 62 clusters) while the voting procedure generates approximately 45 of such clusters (73% of the 62 clusters). This last result is variable due to the randomisation aspect of this approach. At first this might appear as though the voting procedure generates more homogeneous clusters, however it is mainly due to the fact that these are mostly small clusters composed of very few actions. A closer look a the rest of the clusters for both approaches, shows us that the distances can be quite similar (as can be seen in Figure 6).
Similarly, the notion of minimum distance can be used to evaluate the dissimilarity between the profiles for each cluster. For this purpose however, we decided to evaluate the minimum distances only between neighbouring clusters. This gave us the distributions shown in Figure 7.

These last results confirm our understanding of these approaches. The intersection operator generates profiles for the resulting clusters that more dissimilar as the algorithm advances through the iterations. This can be seen in Figure 7, where the blue distribution shows us only very high distances because the profiles have very few points in common. The smallest distance obtained is 0.95. In comparison, the voting procedure generates profiles that are still quite similar even after 200 steps. The smallest distance in this
case is 0.38. We have to remember however that these very different values are due to the way these profiles are constructed.

**Figure 7** Dissimilarity: distribution of the minimum distance between all neighbouring clusters for both approaches (see online version for colours)

![Graph showing dissimilarity distribution](image)

**Figure 8** Distribution of cluster sizes for both approaches (see online version for colours)

![Graph showing cluster size distribution](image)

Finally another way to compare the two approaches would be to analyse the distribution of cluster sizes that we obtain. Figure 8 shows us these two distributions and we can immediately notice that the voting procedure generates 46 clusters (out of 62) that each contain only one municipality along with a single one (not represented in this figure) that contains 152 municipalities. The intersection approach however allows us to obtain a slightly more uniform distribution.
4 Conclusions

In this paper we developed an extension of an existing multicriteria clustering method. We adapted it to the spatial problem of territory partitioning and proposed two variants to the algorithm. The main advantage of this method compared to other districting techniques is that it take into account the multicriteria information of all the areas before grouping them in clusters.

The first difference we introduced was changing the method into a hierarchical clustering technique. This eliminated the problem of initial conditions that influence the results and lead us to a method that can more easily produce stable results.

Out of the two variants we proposed, the first one uses a voting procedure similar to the one in the existing clustering method. The effect of this approach is that all the areas that are similar are more easily grouped while only outliers and particular cases are kept isolated until the last iterations of the algorithm. This is due to the fact that a new artificial profile is built for each cluster of areas which ignores differences that can exist within the cluster.

The other approach we proposed is more strict when it comes to differences between profiles and will therefore only group areas when these are all similar within the group. This second approach therefore gives us results that are stable and which present clusters of roughly the same size.

Further testing could help us imagine new variants of this clustering method. Moreover, applying this technique to other cases or comparing the obtained results to actual studies of territories could help us understand the characteristics that are highlighted by these variants. Finally, the characterisation of the multicriteria geographical partition quality still has to be further investigated.

References


